

Table 1

Protein SEQ ID No:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	9	1676225	BLADNOT05	1510924F6 (LUNGNOT14), 1676225F6 (BLADNOT05), 1676225H1 (BLADNOT05), 1988463R6 (LUNGAST01), 3143066H1 (HNT2AZS07)
2	10	2057235	BEPINOT01	946212H1 (RATRN02), 989326H1 (LVENNOT03), 2057235H1 (BEPINOT01), 2057235R6 (BEPINOT01), 2059357T6 (OVARNOT03), 2885067H1 (SINJNOT02), 4331146H1 (KIDNNOT32)
3	11	2186363	PROSNOT26	029201R1 (SPLNFET01), 350269H1 (LVENNOT01), 982451T2 (TONGTUT01), 2186363H1 (PROSNOT26), 2186363X11C1 (PROSNOT26), 3011602H1 (MUSCNOT07), 4603644H1 (BRSTNOT07)
4	12	3126833	LUNGUTUT12	1212133H1 (BRSTTUT01), 1503929F1 (BRAITUT07), 3126833H1 (LUNGUTUT12)
5	14	3446038	FIBPNOT01	1695215H1 (COLNNOT23), 2103116T6 (BRAITUT02), 2778432F6 (OVARUTUT03), 2969365F6 (HEAONOT02), 3446038H1 (FIBPNOT01), 4462290H1 (HEAADIT01), SBHA01009F1
6	14	4113161	UTRSTUT07	865350R1 (BRAITUT03), 1323856F1 (LPARNOT02), 1573915X11 (LNODNOT03), 1573915X13 (LNODNOT03), 1711996F6 (PROSNOT16), 2432241H1 (BRAVUNT02), 4113161H1 (UTRSTUT07)
7	15	4408678	OVARNOT13	4408537H1 (OVARNOT13), 4408678H1 (OVARNOT13)
8	16	4942111	BRAIFEN03	4942111F6 (BRAIFEN03), 4942111H1 (BRAIFEN03)

PF-0678 US

Table 2

Protein SEQ ID No:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	213	S78 S58 S57 S113 T127		Respiratory-chain NADH dehydrogenase 20 kD subunit: V91-E204	PSST subunit of the NADH:ubiquinone oxidoreductase complex [Bos taurus] g599691	Motifs BLAST BLOCKS Pfam
2	303	S115 T124 T214 S264 T75 S98 T145	N277	Mitochondrial carrier protein: A133-L187	Mitochondrial carrier protein [Homo sapiens] g5815345	Motifs BLAST BLOCKS Pfam
3	294	S164 T195 S229 T183 T231 Y284	N27 N137		ABC1 protein [Arabidopsis thaliana] g3859609	Motifs BLAST
4	179	S105 S170		Transmembrane domain: S20-L37	Divalent cation tolerance protein [Homo sapiens] g4454995	Motifs BLAST HMMER
5	314	S303 T92 T97 T109 T261 S279 S298 S303		Mitochondrial carrier protein: V33-V101, N108-K255, N256-L302	Putative mitochondrial carrier protein [C. elegans] g3879122	Motifs BLAST PRINTS ProfileScan Pfam
6	544	S126 S161 S209 T383 T427 T535 S539 T70 S76 T148 T318 T462		Protein intergenic region, ABC1 precursor, mitochondrial energy transfer: V149-Q158, V207-G226, G231-V251, V309-G329	ABC1 protein [C. elegans] g3859609	Motifs BLOCKS
7	128	S26 S33 T39 Y43	N122		Photosystem II 10 kD polypeptide [Oryza sativa] g1835731	Motifs BLAST
8	72	T31 S69			Ubiquinol-cytochrome c reductase [Solanum tuberosum] g633687	Motifs BLAST

Table 3

Nucleotide SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
9	Nervous (0.186) Reproductive (0.172) Cardiovascular (0.145)	Cell Proliferation (0.669) Inflammation/Immune (0.289)	pINCY
10	Reproductive (0.219) Gastrointestinal (0.190) Nervous (0.143)	Cell Proliferation (0.743) Inflammation/Immune (0.286)	PSPORT1
11	Reproductive (0.256) Nervous (0.140) Gastrointestinal (0.128)	Cell Proliferation (0.686) Inflammation/Immune (0.314)	pINCY
12	Reproductive (0.259) Gastrointestinal (0.165) Nervous (0.147)	Cell Proliferation (0.697) Inflammation/Immune (0.295)	pINCY
13	Reproductive (0.250) Hematopoietic/Immune (0.188) Cardiovascular (0.125)	Cell Proliferation (0.563) Inflammation/Immune (0.437)	pINCY
14	Nervous (0.192) Gastrointestinal (0.154) Reproductive (0.154)	Cell Proliferation (0.557) Inflammation/Immune (0.365)	pINCY
15	Dermatologic (0.333) Cardiovascular (0.333) Reproductive (0.333)	Cell Proliferation (0.666)	pINCY
16	Nervous (1.000)	Cell Proliferation (1.000)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Comment
9	BLADNOT05	Library was constructed using RNA isolated from bladder tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology for the associated tumor tissue indicated grade 3 transitional cell carcinoma. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use.
10	BEPINOT01	Library was constructed using RNA isolated from a bronchial epithelium primary cell line derived from a 54-year-old Caucasian male.
11	PROSNOT26	Library was constructed using RNA isolated from prostate tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology for the associated tumor tissue indicated an adenocarcinoma. The patient presented with elevated prostate specific antigen (PSA). Family history included a malignant stomach neoplasm.
12	LUNGTTUT12	Library was constructed using RNA isolated from tumorous lung tissue removed from a 70-year-old Caucasian female during a lung lobectomy of the left upper lobe. Pathology indicated grade 3 (of 4) adenocarcinoma and vascular invasion. Patient history included tobacco abuse, depressive disorder, anxiety state, and skin cancer. Family history included cerebrovascular disease, congestive heart failure, colon cancer, depressive disorder, and primary liver.
13	FIBPNOT01	Library was constructed using RNA isolated from fibroblasts of the prostate stroma removed from a male fetus, who died after 26 weeks' gestation.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
14	UTRSTUT07	Library was constructed using RNA isolated from uterine tumor tissue removed from a 41-year-old Caucasian female during total abdominal hysterectomy with removal of an ovary and incidental appendectomy. Pathology indicated the endometrium was secretory phase, and the cervix showed microglandular hyperplasia. There were multiple (2 subserosal, 13 intramural, 1 submucosal) leiomyomas. Family history included atherosclerotic coronary artery disease, benign hypertension, depression, and type II diabetes.
15	OVARNOT13	Library was constructed using RNA isolated from left ovary tissue removed from a 47-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy, and dilation and curettage. Pathology for the associated tumor tissue indicated a single intramural leiomyoma. The endometrium was in the secretory phase. The patient presented with menorrhagia. Patient history included hyperlipidemia and benign hypertension. Family history included colon cancer, benign hypertension, atherosclerotic coronary artery disease, and breast cancer.
16	BRAIFEN03	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from the BRAIFET02 library. Starting RNA was made from brain tissue removed from a Caucasian male fetus with a hypoplastic left heart stillborn after 23 weeks' gestation. The library was normalized in two rounds (with 48 hour reannealing hybridizations) using conditions adapted from Soares et al. (Proc. Natl. Acad. Sci. USA (1994) 91:9928) and Bonaldo et al. (Genome Research (1996) 6:791).

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <i>supra</i> ; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	